

Weekly Report

June 17th, 2018

Done:

1. Finish the last class of RNN. All the courses of deeplearning.ai are finished. I've learned the fundamental models of DNN, their origins/theories behind algorithms, and basic coding skills using keras or tensorflow.
2. A discuss on the data/vis provenance topic, several key points:
 - (1) visually organize what the users have done
 - (2) design should be simple
 - (3) define metrics to weigh and visual the task/user operation saliency
 - (4) consider other applications, not only about fire data system
3. (1) Learn to use the weather API by Max. He split the previous routers, and I will adapt it to the system. (2) Adapt the area selection in the system.

To Do:

1. We will have a meeting with emergency office.
2. A draft of prototype.
3. More progress of the system: hover and show the weather information

Paper reading:

1. TKDE05 *Toward Intelligent Assistance for a Data Mining Process: An Ontology-Based Approach for Cost-Sensitive Classification* 对于数据挖掘过程的智能支持, 他会根据数据特征和想要的结果, 用 ontology 的方法来搜索可用的数据挖掘过程. 此外还通过一定准则来 ranking 这些过程, 帮助作出推荐.

2. *Provenance for Data Mining* 文章阐述了对数据挖掘过程做 provenance 的重要性. 文中提出了多种方法, 比如输入如何影响结果, 这种关系是怎样的, (如何)提供上下文信息. 从一定角度来看, 他很像是一些开黑盒方法: 哪个 input 会造成什么结果, 没有他会怎样, 或者没有了一个中间介质再没有这个 input 会怎样. 这个思路又有点像上周那篇 CHI 的交互式 DR 的文章, 但是本文发表于一个信息安全会议

3. EuroVis18 *Chart Constellations: Effective Chart Summarization for Collaborative and Multi-User Analyses* 这边小组的一个作品, 其侧重点是协同. 背景是一帮用户分析一个数据集, 然后用的是基本的可视化方法来发现其中的关系. 这些可视化彼此之间有一定关联, 他们定义一些距离矩阵来做一个投影(MetaVis, vis of vis); 在此基础上可以进去不同的可视化看, 可以改, 可以画个新的, 可以发现哪些还没探索等等. 其实我们可以不一定用在用户协同领域, 对平时分析也有不小的意义.

4. CHI17 *GraphScape: A model for automated reasoning about visualization similarity and sequencing* 这也是上文中用到的一种度量方法, 用了一个有向图来对可视化设计空间进行建模, 支持定于可视化相似性和顺序性的推理. 它是基于 vega-lite 的, 据我最近观察现在也有不少 vega-lite(语法)类似物, 所以这个方法应会有普适性.

5 Word2Vec: 这个应用很常见, 但是理解其算法仍然有点难度. 本身是训练了一个简单的

只有一个隐藏层的 NN，但是其本意并不是为了用这个 NN，而是用它隐藏层的系数，作为 embedding. 为什么这样做，有个简单的例子，我要喝苹果*；我要喝梨*，我们知道*代汁，用 one-hot 是无法体现的，只有通过隐藏层的乘积再进行一系列的变化才能学到这个水果后面加“汁”的关系. 这个隐藏层就可以当做 embedding；推广一下就是 skip-gram；然后为了降低复杂度，用了 negative sampling.